

Checking Tests for Read-Once Functions over Arbitrary Bases

Dmitry V. Chistikov

Faculty of Computational Mathematics and Cybernetics
Moscow State University, Russia
dch@cs.msu.ru

Abstract. A Boolean function is called read-once over a basis B if it can be expressed by a formula over B where no variable appears more than once. A checking test for a read-once function f over B depending on all its variables is a set of input vectors distinguishing f from all other read-once functions of the same variables. We show that every read-once function f over B has a checking test containing $O(n^l)$ vectors, where n is the number of relevant variables of f and l is the largest arity of functions in B . For some functions, this bound cannot be improved by more than a constant factor. The employed technique involves reconstructing f from its l -variable projections and provides a stronger form of Kuznetsov's classic theorem on read-once representations.

Keywords: read-once Boolean function, checking test, complexity, teaching dimension, equivalence query, membership query.

1 Introduction

Let B be an arbitrary set of Boolean functions. A function f is called *read-once* over B iff it can be expressed by a formula over B where no variable appears more than once.

Let $f(x_1, \dots, x_n)$ be a read-once function over B that depends on all its variables. Then a set M of n -bit vectors is called a *checking test* for f iff for any other read-once function $g(x_1, \dots, x_n)$ over B there exists a vector $\alpha \in M$ such that $f(\alpha) \neq g(\alpha)$. In other words, M is a checking test for f iff the values of f on vectors from M distinguish f from all other read-once functions g of the same variables. Note that all these *alternatives* g , unlike the *target function* f , may have irrelevant variables.

Denote by B_l the basis of all l -variable Boolean functions. The goal of this paper is to prove that all n -variable read-once functions over B_l have checking tests containing $O(n^l)$ vectors. More generally, for an arbitrary basis B and a read-once function f over B , denote by $T_B(f)$ the smallest possible number of vectors in a checking test for f . This value can be regarded as the *checking test complexity* of f . We show that

$$T_{B_l}(f(x_1, \dots, x_n)) \leq 2^l \cdot \binom{n}{l}$$

and, therefore, for any finite basis B and any sequence of read-once functions f_n of n variables over B , it holds that $T_B(f_n) = O(n^l)$ as $n \rightarrow \infty$, where l is the largest arity of functions from B .

This result is based on the previously known relevance hypercube method by Voronenko [15]. Our main contribution is the proof that the method is *correct* for all bases B_l for an arbitrary l , i. e., it provides a way to construct checking tests of specified *length* (cardinality) for all read-once functions over these bases. Previous results give proofs only for $l \leq 5$ [14,15,18].

It should be pointed out that the bound $T_{B_l}(f) = O(n^{l+1})$ can be extracted from the related paper on the exact identification problem by Bshouty, Hancock and Hellerstein [4]. Our result has the following advantages. Firstly, for some functions the bound $O(n^l)$ cannot be improved by more than a constant factor (it matches the known lower bound $\Omega(n^l)$ for n -ary disjunction up to a constant factor). Secondly, checking tests constructed by the relevance hypercube method have regular structure. In short, we show that every read-once function can be unambiguously reconstructed from a set of its l -variable projections with certain properties. This fact may look natural at first sight, but turns out a tricky thing to prove after all.

2 Background and Related Work

Our result has some interesting consequences related to computational learning theory. For instance, it is known that checking tests can be used to implement *equivalence queries* from Angluin's learning model [2]. For the problem of identifying an unknown read-once function over an arbitrary finite basis B with queries, it turns out that non-standard *subcube identity queries* can be efficiently used [7]. A subcube identity query basically asks whether a specified projection of the unknown function f is constant, i. e., whether a given partial assignment of constants to input variables unambiguously determines the value of f .

It follows from our results that for any finite basis B , the problem of learning an unknown read-once function over B can be solved by an algorithm making $O(n^{l+2})$ membership and subcube identity queries, which is polynomial in n (here l is the largest arity of functions in B and a standard *membership query* is simply a request for the value of f on a given input vector). This result builds upon an algorithm by Bshouty, Hancock and Hellerstein [4], which is a strong generalization of a classic exact identification algorithm by Angluin, Hellerstein and Karpinski [3].

Closely related to the notion of checking test complexity is the definition of teaching dimension introduced by Goldman and Kearns [12]. A *teaching sequence* for a Boolean concept (a Boolean function f) in a known class \mathcal{C} is a sequence of labeled instances (pairs of the form $\langle \alpha, f(\alpha) \rangle$) consistent with only one function f from \mathcal{C} . The *teaching dimension* of a class is the smallest number t such that all concepts in the class have teaching sequences of length at most t .

In test theory, which dates back to 1950s [6], the corresponding definition is that of the *Shannon function for test complexity*, which is the largest test com-

plexity of an n -variable function. In these terms, our Corollary can be restated as follows: $T_{B_l}(n) = O(n^l)$, where $T_{B_l}(n)$ is the Shannon function for checking test complexity of read-once Boolean functions (i.e., the maximum of $T_{B_l}(f)$ over all n -variable read-once functions over B_l). It must be stressed that in the definition of a checking test used in this paper, the target function is required to depend on all its variables. (One needs to test all 2^n input vectors to distinguish the Boolean constant 0 from all read-once conjunctions of n literals.)

Another appealing problem is that of obtaining bounds on the value of $T_B(f)$ for individual read-once functions f . The bound $T_{B_l}(x_1 \vee \dots \vee x_n) = \Theta(n^l)$ is obtained in [15] and generalized in [16]. In [8], it is shown that for a wide class of bases including B_l , $l \geq 2$, there exist pairs of read-once functions f, f' such that f' is obtained from f by substituting a constant for a variable and $T_B(f') > T_B(f)$. This result shows that lower bounds on $T_B(f)$ cannot generally be obtained by simply finding projections of f that are already known to require a large number of vectors in their checking tests.

For the basis B_2 , individual bounds on the checking test complexity are obtained in [17]. In [19], it is shown that almost all read-once functions over the basis $\{\vee, \oplus\}$ have a relatively small checking test complexity of $O(n \log n)$, as compared to the maximum of $\Theta(n^2)$ (even if alternatives are arbitrary read-once functions over B_2 and not necessarily read-once over $\{\vee, \oplus\}$). For the standard basis $\{\wedge, \vee, \neg\}$, it is known that $n+1 \leq T_{\{\wedge, \vee, \neg\}}(n) \leq 2n+1$ [10], and individual bounds can be deduced from those for the monotone basis $\{\wedge, \vee\}$ [5,9].

3 Basic Definitions

A variable x_i of a Boolean function $f(x_1, \dots, x_n)$ is called *relevant* (or *essential*) if there exist two n -bit vectors α and β differing only in the i th component such that $f(\alpha) \neq f(\beta)$. If x_i is relevant to f , then f is said to *depend* on x_i .

In this paper, we call a pair of functions $f(x_1, \dots, x_n)$ and $g(y_1, \dots, y_n)$ *similar* if for some constants $\sigma, \sigma_1, \dots, \sigma_n \in \{0, 1\}$ and for some permutation π of $\{1, \dots, n\}$ the following equality holds:

$$f(x_1, \dots, x_n) \equiv g^\sigma(x_{\pi(1)}^{\sigma_1}, \dots, x_{\pi(n)}^{\sigma_n}),$$

where z^τ stands for z if $\tau = 1$ and for \bar{z} if $\tau = 0$. A Boolean function $f(x_1, \dots, x_n)$, $n \geq 3$, is called *prime* if it has no decomposition of the form

$$f(x_1, \dots, x_n) \equiv g(h(x_{\pi(1)}, \dots, x_{\pi(k)}), x_{\pi(k+1)}, \dots, x_{\pi(n)}),$$

where $1 < k < n$ and π is a permutation of $\{1, \dots, n\}$.

The structure of formulae expressing read-once functions can be represented by rooted trees. A *tree* of a read-once function $f(x_1, \dots, x_n)$ over B_l has n leaves labeled with literals of different variables and one or more internal nodes labeled with functions from B_l and symbols $\circ \in \{\wedge, \vee, \oplus\}$ of arbitrary arity (possibly exceeding l). We assume without loss of generality that such trees also have the following properties:

- 1) any internal node is labeled either with a prime function or with a symbol $\circ \in \{\wedge, \vee, \oplus\}$;
- 2) internal nodes labeled with identical symbols $\circ \in \{\wedge, \vee, \oplus\}$ are not adjacent.

One can readily see that every read-once function over B_l has at least one tree of this form.

In the sequel, variables are usually identified with corresponding leaves in the tree. Denote by $\text{lca}(y_1, \dots, y_m)$ the *least common ancestor* of variables y_1, \dots, y_m , i.e., the last common node in (simple) paths from the root of the tree to y_1, \dots, y_m .

Suppose that T is a tree of a read-once function and v is its internal node. By T_v we denote the *subtree of T rooted at v* , i.e., the rooted tree that has root v and contains all descendants of v . If w_1, \dots, w_p are *children* (direct descendants) of v , then subtrees T_{w_1}, \dots, T_{w_p} are called *subtrees of the node v* . If x is a leaf of T contained in T_v , then by T_v^x we denote a (unique) subtree T_{w_j} containing x . Finally, subtrees of the root node of a tree are called *root subtrees*.

4 The Relevance Hypercube Method

This section is devoted to the review of the relevance hypercube method proposed by Voronenko in [15]. This method has been known to be correct for the bases B_l if $l \leq 5$ (see [15,18]).

From now on, we will use the term “read-once function” instead of “read-once function over B_l ”. We use boldface letters to denote vectors (often treated as sets) of variables.

Let f be a read-once function depending on variables $\mathbf{x} = \{x_1, \dots, x_n\}$. Suppose that H is a set of 2^l input vectors disagreeing at most in i_1 th, \dots , i_l th components such that the restriction of f to H (which is an l -variable Boolean function) depends on all its l variables $\mathbf{x}' = \{x_{i_1}, \dots, x_{i_l}\}$. Then H is called a *relevance hypercube* (or an *essentiality hypercube*) of dimension l for these variables \mathbf{x}' . Any relevance hypercube can be identified with a partial assignment p of constants to input variables such that the induced projection f_p depends on all its l variables. Such assignments are called l -justifying in [4].

Remark. For some functions f and some subsets of their variables relevance hypercubes do not exist. For instance, one may easily check that the function $d(x, u_0, u_1) = (\bar{x} \wedge u_0) \vee (x \wedge u_1)$ has no relevance hypercubes for the set $\mathbf{u} = \{u_0, u_1\}$. As indicated below, the absence of relevance hypercubes is a major obstacle to proving the correctness of the relevance hypercube method (see also [15,18]). At the same time, for some functions there exist subsets of variables with more than one relevance hypercube. An example is given by the same function d and the set $\mathbf{u}' = \{x, u_0\}$.

Any set M of input vectors is called a *relevance hypercube set* of dimension l for f if it contains a relevance hypercube H for every l -sized subset of \mathbf{x} , for which such a hypercube exists. (Recall that \mathbf{x} is the set of all variables relevant to

f .) In other words: consider all l -sized subsets $\mathbf{w} \subseteq \mathbf{x}$ such that f has a relevance hypercube for \mathbf{w} . A set M is a relevance hypercube set iff M contains at least one relevance hypercube for each subset \mathbf{w} of this kind. It was conjectured that any such set is a checking test for f .

Suppose that f is a read-once function that depends on n variables \mathbf{x} . Construct a *relevance table* with $\binom{n}{l}$ rows and two columns by the following rule. First, fill the first cells of all rows with different l -sized subsets of \mathbf{x} . Then for each row, if the first cell contains a subset $\mathbf{w} \subseteq \mathbf{x}$, put in the second cell any relevance hypercube for \mathbf{w} (along with the corresponding values of f) if such a hypercube exists, or the symbol $*$ otherwise.

In [15], it is shown that any (valid) relevance table uniquely determines a read-once function in the following sense. Suppose that one knows that a function g is read-once and agrees with f on all relevance hypercubes from a relevance table E for f . If one also knows that for each $*$ -row in E the function g does not have a relevance hypercube, then one can recursively reconstruct the *skeleton* of g (which is a tree T' such that negating some of its nodes' labels yields a correct tree representing g). After that, the values of f on the vectors from E allow one to prove that g is equal to f .

It follows that a relevance hypercube set M of dimension l for a read-once function f is indeed a checking test for f if E contains no $*$ -rows. If for some l -sized subset of variables \mathbf{w} no relevance hypercube exists, then a more sophisticated technique is needed to prove that f can still be reconstructed from its values on vectors from M . The approach used in this paper is outlined in the following Section 5.

5 Assumptions and Notation

In this section we make preliminary assumptions and introduce some notation. All subsequent work, including the proof of our main theorem, is based on the material presented here.

We start with an arbitrary read-once function f over B_l , where $l \geq 3$. Let \mathbf{x} be the set of variables relevant to f . Suppose that M is a relevance hypercube set of dimension l for f . Our goal is to prove that M is a checking test for f , i. e., for any other read-once function $g(\mathbf{x})$ there exists a vector $\alpha \in M$ such that $f(\alpha) \neq g(\alpha)$.

Take any read-once function $g(\mathbf{x})$ that agrees with $f(\mathbf{x})$ on all vectors from M . Firstly and most importantly, we need to prove that root nodes of these two functions' trees are labeled with similar functions. If either of the root nodes is labeled with a symbol $\circ \in \{\wedge, \vee, \oplus\}$, then this can be done with the aid of techniques similar to those from [15]. Here we focus on the prime case, i. e., we assume that

$$\begin{aligned} f(\mathbf{x}) &= f^0(f_1(\mathbf{x}^1), \dots, f_s(\mathbf{x}^s)), \\ g(\mathbf{y}) &= g^0(g_1(\mathbf{y}^1), \dots, g_r(\mathbf{y}^r)), \end{aligned}$$

where both f^0 and g^0 are prime, and $\mathbf{x}^1 \cup \dots \cup \mathbf{x}^s$ and $\mathbf{y}^1 \cup \dots \cup \mathbf{y}^r$ are partitions of $\mathbf{x} = \mathbf{y}$. (Technically, we must first assume that $\mathbf{y} \subseteq \mathbf{x}$, but it is easily shown that no variable from \mathbf{x} can be irrelevant to g ; see, e.g., Proposition 2 in the next section.) Note that here $s \leq l$ and $r \leq l$.

Suppose we have already proved that f^0 and g^0 are similar. As our second step, we need to show that partitions of input variables into subtrees are identical in the representations above. In other words, we need to show that each \mathbf{y}^k is equal to some \mathbf{x}^i .

These two steps, especially the first one, constitute the main difficulties in proving the correctness of the method. The remaining part is technical and can be done with the aid of induction on the depth of the tree representing f . A short explanation of how this part is done is given at the end of our main theorem's proof in Section 8.

In the following sections, we will need the *colouring* of input variables \mathbf{x} defined by the following rule. To each variable $x \in \mathbf{x}$ we assign a (unique) colour $k \in \{1, \dots, r\}$ such that $x \in \mathbf{y}^k$. This definition provides a convenient way of relating functions f and g (i.e., their tree structure) to each other.

6 Some Observations

In this section we present three facts needed for the sequel. A key observation is given by the following proposition.

Proposition 1. *Suppose that g' is a projection of g that depends on variables x and y having the same colour k . Also suppose that $\text{lca}(x, y) = v$ in a tree T' of g' . Then, if v is labeled with a prime function h , it follows that all leaves in the subtree $(T')_v$ have the same colour k . Otherwise, if v is labeled with a symbol $\circ \in \{\wedge, \vee, \oplus\}$, it follows that all leaves in subtrees $(T')_v^x$ and $(T')_v^y$ have the same colour k .*

Proposition 1 follows from a simple fact that substitutions of constants for variables of g can result in removing nodes and subtrees from T' , or in replacing nodes with trees that represent projections of prime functions. Adjacent nodes labeled with identical symbols $\circ \in \{\wedge, \vee, \oplus\}$ are subsequently glued together, but least common ancestors of each \mathbf{y}^i either remain roots of single-coloured subtrees, or “support” subsets of single-coloured subtrees of internal nodes labeled with $\circ \in \{\wedge, \vee, \oplus\}$.

For technical reasons, we will also need the following proposition, which holds true for all discrete functions (not necessarily read-once or even Boolean) and follows from Theorem B in [11].

Proposition 2. *Suppose that f is an arbitrary function depending on n variables \mathbf{x} . Also suppose that there exists a relevance hypercube for some p variables $\mathbf{u} \subseteq \mathbf{x}$. Then for every q such that $p \leq q \leq n$ there exists a relevance hypercube for some q -sized set of variables \mathbf{w} such that $\mathbf{u} \subseteq \mathbf{w} \subseteq \mathbf{x}$.*

Last but not least, we will use the following fact (see, e.g., [15]).

Proposition 3. *Suppose that a read-once function f is represented by a tree and p variables \mathbf{u} are taken from p different subtrees of an internal node v , which is labeled with a prime function of arity p or with a symbol $\circ \in \{\wedge, \vee, \oplus\}$. Then f has at least one relevance hypercube for \mathbf{u} and, moreover, restrictions of f to all such hypercubes are:*

- (a) *similar to $h(z_1, \dots, z_p)$ if v is labeled with a prime function h ;*
- (b) *similar to $z_1 \circ \dots \circ z_p$ if v is labeled with a symbol $\circ \in \{\wedge, \vee, \oplus\}$.*

7 Auxiliary Lemmas

Suppose that read-once functions f and g satisfy all the assumptions made in Section 5 and T is a tree of f . Recall that by \mathbf{x} we denote the set of all variables relevant to f . We say that a set $\mathbf{u} \subseteq \mathbf{x}$ is *stable* iff for any set \mathbf{w} such that $\mathbf{u} \subseteq \mathbf{w} \subseteq \mathbf{x}$ and any relevance hypercube H for \mathbf{w} there exists a relevance hypercube H' for \mathbf{u} such that $H' \subseteq H$.

Remark. The definition of a stable set \mathbf{u} does not require the existence of relevance hypercubes for all sets \mathbf{w} such that $\mathbf{u} \subseteq \mathbf{w} \subseteq \mathbf{x}$. What it says is that if there exists such a relevance hypercube H , then there exists a relevance hypercube for \mathbf{u} which is a subcube of H . For $\mathbf{w} = \mathbf{x}$, however, the definition requires that at least one relevance hypercube for \mathbf{u} exists.

One can readily observe that all singleton subsets of \mathbf{x} are stable. Examples of sets that are not stable are given by $\mathbf{u} = \{u_0, u_1\}$ (as witnessed by $\mathbf{w} = \mathbf{u} \cup \{y\}$) for functions $f_1 = (\bar{x} \wedge d(y, u_0, u_1)) \vee (x \wedge (y \vee u_0 \vee u_1))$ and $f_2 = (\bar{x} \wedge d(y, u_0, u_1)) \vee (x \wedge (u_0 \vee u_1))$, where $d(y, u_0, u_1) = (\bar{y} \wedge u_0) \vee (y \wedge u_1)$. Our main ingredient in the proof of the main theorem is given by the following lemma, which allows us to establish a link between the tree structure of our functions f and g .

Lemma 1. *For any stable set \mathbf{u} of at most l variables of the function f , the function g agrees with f on some relevance hypercube for \mathbf{u} .*

Proof. We start with the definition of a stable set. First choose $\mathbf{w} = \mathbf{x}$ and conclude that f has a relevance hypercube for \mathbf{u} . By Proposition 2, f also has a relevance hypercube for some l -sized set of variables \mathbf{w}' such that $\mathbf{u} \subseteq \mathbf{w}'$. It follows that M contains some relevance hypercube for \mathbf{w}' . Since f and g agree on all vectors from M , and \mathbf{u} is stable, it also follows that f and g agree on some relevance hypercube for \mathbf{u} . This concludes the proof.

More than once we will need subsets of input variables having specific structure. We call a set \mathbf{u} *conservative* iff for each internal node v of T labeled with a prime function h the number of subtrees of v containing at least one variable from \mathbf{u} is equal either to 0, or to 1, or to the arity of h . To understand the intuition behind this term, consider the restriction of f to any relevance hypercube for such a set. In the tree of such a restriction, each node of T labeled with a prime function is either preserved or discarded, i.e., no constant substitutions and further transformations occur at these nodes.

Lemma 2. *All conservative sets are stable.*

Proof. Let \mathbf{u} be a conservative set of variables. We say that an internal node of T is a *branching node* for \mathbf{u} iff at least two subtrees of v contain leaves from \mathbf{u} . The proof is by induction over the number b of branching nodes for \mathbf{u} in T . If $b = 0$, then $|\mathbf{u}| \leq 1$ and there is nothing to prove. Suppose that $b \geq 1$ and $v = \text{lca}(\mathbf{u})$. Then v is a branching node and all other branching nodes are descendants of v . Therefore, f can be expressed by a formula

$$h_0(\mathbf{z}^0, h(h_1(\mathbf{z}^1), \dots, h_m(\mathbf{z}^m))),$$

where h is the function corresponding to the label of v , sets \mathbf{z}^i and \mathbf{z}^j are disjoint for $i \neq j$ and (by our definition of a conservative set) variables \mathbf{u} are contained in each \mathbf{z}^i , $i \geq 1$, but not in \mathbf{z}^0 .

Identify a relevance hypercube H for some variables \mathbf{w} (here $\mathbf{u} \subseteq \mathbf{w}$) with a partial assignment p of constants to variables \mathbf{x} . Split p into p_0, p_1, \dots, p_m according to the partitioning given by $\mathbf{z}^0, \mathbf{z}^1, \dots, \mathbf{z}^m$. Subsets of \mathbf{u} contained in \mathbf{z}^i , $i \geq 1$, are conservative for trees representing functions $h_i(\mathbf{z}^i)$, and the number of branching nodes in any such tree is at most $b - 1$. By the inductive assumption, there exist partial assignments p'_1, \dots, p'_m which are extensions of p_1, \dots, p_m and restrict relevance hypercubes for these subsets. Projections h'_i induced by p'_i depend on these subsets of \mathbf{u} . If we now choose an extension p'_0 of p_0 taking $h_0(\mathbf{z}^0, u)$ to a literal of u , the composition of p'_0, p'_1, \dots, p'_m will restrict the needed relevance hypercube H' . This concludes the proof.

In Section 5, we defined the colouring of variables induced by the read-once representation of g . The following lemmas reveal some properties of this colouring that are related to the structure of T . These properties reflect the observation formulated in Proposition 1.

Lemma 3. *Suppose that variables x and y both have colour k . Also suppose that the node $v = \text{lca}(x, y)$ in T is labeled with a prime function h . Then all the leaves in the subtree T_v have the same colour k .*

Proof. Let m be the arity of h . Take arbitrary variables z_1, \dots, z_{m-2} such that $x, y, z_1, \dots, z_{m-2}$ are contained in m different subtrees of v . The set $\mathbf{u} = \{x, y, z_1, \dots, z_{m-2}\}$ is conservative and, therefore, stable (by Lemma 2). Since h is prime and f is read-once over B_l , we see that $m \leq l$. It follows from Lemma 1 that f agrees with g on some relevance hypercube for \mathbf{u} . By Proposition 3, the restriction of f to any such hypercube is similar to h . Therefore, some projection g' of g is represented by a tree T' with exactly one internal node, which is labeled with a prime function. Variables x and y are relevant to g' and have the same colour k . It then follows from Proposition 1 that all variables z_1, \dots, z_{m-2} have colour k too. Since z_1, \dots, z_{m-2} were chosen arbitrarily from their subtrees, we obtain that all leaves in $T_v^{z_1}, \dots, T_v^{z_{m-2}}$ have colour k . Repeating the same reasoning for initial pairs x, z_1 and y, z_1 in place of x, y reveals that all leaves in T_v^y and T_v^x also have the same colour k . This concludes the proof.

Lemma 4. *For each colour $k \in \{1, \dots, r\}$, there exists a unique index $i \in \{1, \dots, s\}$ such that $\mathbf{y}^k \subseteq \mathbf{x}^i$, i. e., all variables coloured with k belong to the set \mathbf{x}^i .*

Proof. Take any two variables x and y having colour k . If they do not belong to the same \mathbf{x}^i , then they belong to different root subtrees of T . Therefore, the node $\text{lca}(x, y)$ is labeled with a prime function f^0 . By Lemma 3, all leaves of T have the same colour, which is a contradiction.

Another way to state Lemma 4 is to say that the partition $\mathbf{y}^1 \cup \dots \cup \mathbf{y}^r$ is a refinement of $\mathbf{x}^1 \cup \dots \cup \mathbf{x}^s$.

Lemma 5. *For any non-root internal node v in T labeled with a prime function h of arity r or greater, all leaves of T_v have the same colour.*

Proof. Let v be an internal node of T labeled with a prime function h of arity at least r . If not all leaves of T_v have the same colour, then by Lemma 3 any two leaves from different subtrees of v have different colours. It then follows that leaves of T_v are coloured with at least r colours. By Lemma 4, leaves of other root subtrees of T cannot be coloured, since all colours are taken from the set $\{1, \dots, r\}$. This contradiction concludes the proof.

Lemma 6. *Suppose that variables x and y both have colour k . Also suppose that in T the node $v = \text{lca}(x, y)$ is labeled with a symbol $\circ \in \{\wedge, \vee, \oplus\}$. Then all the leaves in T_v^x and T_v^y have the same colour k .*

Proof. Define the *depth* of a subtree as the maximum number of edges on (shortest) paths from its root to its leaves. Let d be the depth of T_v . The proof is by induction over d . For $d = 1$, there is nothing to prove. Suppose that $d \geq 2$ and T_v^y contains a leaf z having colour $k' \neq k$. We claim that for some $m \geq 0$ there exist variables z_0, z_1, \dots, z_m such that the set $\mathbf{u} = \{x, y, z_0, z_1, \dots, z_m\}$ is conservative, has cardinality at most l , and the restriction of f to any relevance hypercube for \mathbf{u} can be obtained by negating the inputs and/or the output of some function $x \circ f'(y, z_0, z_1, \dots, z_m)$, where the colours of y and z_0 are different and f' is either a prime function or a binary function ($m = 0$) from $\{\wedge, \vee, \oplus\}$ different from \circ .

First suppose that the root w of T_v^y is labeled with a prime function h . Since not all leaves of T_v^y have the same colour, it follows from Lemma 3 that colours of leaves taken from different subtrees of T_v^y are different. Take arbitrary variables z_0, z_1, \dots, z_m , $m \geq 1$, from all subtrees except T_w^y (one variable from each subtree). Now y and z_0 have different colours and $m+2 \leq r-1$ by Lemma 5. One can easily see that the set \mathbf{u} constructed in this way is conservative by definition and has cardinality at most l , because $r \leq l$. Proposition 3 then reveals that the restriction of f to any relevance hypercube for \mathbf{u} indeed has the needed form.

Now consider the case when w is labeled with a symbol $\star \in \{\wedge, \vee, \oplus\}$. By definition of a tree representing a read-once function, \star is different from \circ . Observe that the depth of the subtree T_w is less than or equal to $d-1$, so we can

use the inductive assumption for T_w . If y and z belong to the same subtree T' of w , then it follows that only leaves from T' can have the same colour as y . In this case, any leaf from any other subtree can be chosen to be z_0 . In the other case, if y and z belong to different subtrees, simply put $z_0 = z$. One can now see that $m = 0$ and $\mathbf{u} = \{x, y, z_0\}$ satisfy all the stated conditions.

Now apply Lemma 1 to the set \mathbf{u} (recall that all conservative sets are stable by Lemma 2). It follows that g agrees with f on some relevance hypercube for \mathbf{u} . By our choice of \mathbf{u} , this means that g has a projection g' of the form specified above. In the tree of g' , the root is adjacent to the leaf labeled with a literal of x and to the other internal node, whose children are y, z_0, z_1, \dots, z_m . Since x and y have the same colour, it follows from Proposition 1 that z_0 has the same colour as x , which contradicts our choice of z_0 . This completes the proof.

8 Main Theorem

Theorem. *For any read-once function f over B_l , $l \geq 3$, depending on all its variables, any relevance hypercube set of dimension l for f constitutes a checking test for f .*

Proof. Let M be a relevance hypercube set of dimension l for f . By g denote an alternative read-once function that agrees with f on all vectors from M . Suppose that f and g satisfy all the assumptions made in Section 5 and let T be a tree of f . Choose a subset \mathbf{u} of T 's leaves according to the following (non-deterministic) recursive rules:

1. Put $\mathbf{u} = \mathbf{u}(v_0)$, where v_0 is the root of T .
2. If all leaves in T_v have the same colour, then $\mathbf{u}(v) = \{x_i\}$ for some leaf x_i contained in T_v .
3. Otherwise:
 - (a) if v is labeled with a prime function h , then $\mathbf{u}(v) = \bigcup \mathbf{u}(v_i)$ over all children v_i of v ;
 - (b) otherwise, if v is labeled with a symbol $\circ \in \{\wedge, \vee, \oplus\}$, then $\mathbf{u}(v) = \mathbf{u}'(v) \cup \mathbf{u}''(v)$, where $\mathbf{u}'(v) = \bigcup \mathbf{u}(v_i)$ over all multi-coloured subtrees T_{v_i} of v , and $\mathbf{u}''(v) = \bigcup \mathbf{u}(v_j)$ over some subset of all single-coloured subtrees T_{v_j} of v that contains one subtree of each colour.

One can easily see that \mathbf{u} is conservative and, by Lemma 2, stable. By Lemmas 3 and 6, it contains exactly r leaves. It follows from Lemma 1 that g agrees with f on some relevance hypercube H for \mathbf{u} . Since all elements of \mathbf{u} have different colours, it follows from Proposition 3 that the restriction of g to H is similar to g^0 . On the other hand, \mathbf{u} contains at least one leaf from each root subtree of T , so the restriction of f to H has the form

$$f' = f^0(f'_1, \dots, f'_s),$$

where functions f'_i depend on disjoint sets of variables from \mathbf{u} . These two restrictions are equal, so f' is a prime function, which is only possible if $r = s$ and

g^0 is similar to f^0 . By Lemma 4, sets of root subtrees' variables are the same for f and g . This means that

$$g(\mathbf{x}) = f^0(g'_1(\mathbf{x}^1), \dots, g'_s(\mathbf{x}^s)).$$

Since a relevance hypercube set for f contains relevance hypercube sets for all functions $f_1(\mathbf{x}^1), \dots, f_s(\mathbf{x}^s)$ (or their negations) regarded as projections of f , the whole argument can be repeated recursively. In the end, one sees that f and g can be expressed by the same formula, and so $f = g$. This concludes the proof.

Corollary. *For any read-once function f over B_l depending on n variables it holds that*

$$T_{B_l}(f) \leq 2^l \cdot \binom{n}{l} = O(n^l).$$

9 Discussion

It is interesting to note that our result gives a stronger form of Kuznetsov's classic theorem on read-once representations [13]. The original result can be reformulated as follows: for any given Boolean function f and any two trees T_1 and T_2 representing f , there exists a one-to-one correspondence ϕ between the sets of internal nodes of T_1 and T_2 such that the functions represented by each pair of matching nodes are either equal or each other's negations. This fact was independently proved by Aaronson [1], who also developed an $O(N^{\log_2 3} \log N)$ algorithm for transforming the truth table of f into such a tree. Note that a sequence of $O(N)$ -sized circuits that check the existence of and output read-once representations over B_l for any fixed l was constructed in [15]. In these results $N = 2^n$ is the input length.

Now suppose T_1 and T_2 are trees representing n -variable Boolean functions f_1 and f_2 , and it is known a priori that these trees do not contain nodes labeled with prime functions of arity greater than l . Our technique reveals that in order to prove the existence of a correspondence ϕ it is sufficient to verify that f_1 and f_2 agree on an $O(n^l)$ -sized set of input vectors. While Kuznetsov's theorem does not concern itself with computational issues, our theorem shows that only a small fraction of input vectors (in fact, a polynomial number of them, as compared to the total of 2^n) is needed to certify the "similarity" of the trees.

Acknowledgements. The author is indebted to Prof. Andrey A. Voronenko, who suggested the problem considered in this paper. The author also wishes to thank Maksim A. Bashov for useful discussions and the anonymous referees for their advice. This research has been supported by Russian Presidential grant MD-757.2011.9.

References

1. Aaronson, S.: Algorithms for Boolean function query properties. *SIAM Journal on Computing* 32(5), 1140–1157 (2003)
2. Angluin, D.: Queries and concept learning. *Machine Learning* 2(4), 319–342 (1988)
3. Angluin, D., Hellerstein, L., Karpinski, M.: Learning read-once formulas with queries. *Journal of the ACM* 40, 185–210 (1993)
4. Bshouty, N.H., Hancock, T.R., Hellerstein, L.: Learning Boolean read-once formulas over generalized bases. *Journal of Computer and System Sciences* 50(3), 521–542 (1995)
5. Bubnov, S. E., Voronenko, A. A., Chistikov, D. V.: Some test length bounds for non-repeating functions in the $\{\&, \vee\}$ basis. *Computational Mathematics and Modeling* 21(2), 196–205 (2010)
6. Chegis, I. A., Yablonsky, S. V.: Logical methods for controlling electric circuits. *Trudy Matematicheskogo Instituta Steklova* 51, 270–360 (1958) (in Russian)
7. Chistikov, D. V.: On the relationship between diagnostic and checking tests of the read-once functions. *Discrete Mathematics and Applications* 21(2), 203–208 (2011)
8. Chistikov, D. V.: Read-once functions with hard-to-test projections. *Moscow University Computational Mathematics and Cybernetics* 34(4), 188–190 (2010)
9. Chistikov, D. V.: Testing monotone read-once functions. In: Iliopoulos, C. S., Smyth, W. F. (eds.) *IWOCA 2011*. LNCS, vol. 7056, pp. 121–134. Springer, Heidelberg (2011)
10. Chistikov, D. V.: Testing read-once functions over the elementary basis. *Moscow University Computational Mathematics and Cybernetics* 35(4), 189–192 (2011)
11. Davies, R. O.: Two theorems on essential variables. *Journal of the London Mathematical Society* 41(2), 333–335 (1966)
12. Goldman, S. A., Kearns, M. J.: On the complexity of teaching. *Journal of Computer and System Sciences* 50(1), 20–31 (1995)
13. Kuznetsov, A. V.: On read-once switching circuits and read-once compositions of functions in the algebra of logic. *Trudy Matematicheskogo Instituta Steklova* 51, 186–225 (1958) (in Russian)
14. Voronenko, A. A.: On checking tests for read-once functions. In: *Matematicheskie Voprosy Kibernetiki*, vol. 11, pp. 163–176. Fizmatlit, Moscow (2002) (in Russian)
15. Voronenko, A. A.: Recognizing the nonrepeating property in an arbitrary basis. *Computational Mathematics and Modeling* 18(1), 55–65 (2007)
16. Voronenko, A. A.: Testing disjunction as a read-once function in an arbitrary unrepeated basis. *Moscow University Computational Mathematics and Cybernetics* 32(4), 239–240 (2008)
17. Voronenko, A. A., Chistikov, D. V.: Learning read-once functions individually. *Uchenye Zapiski Kazanskogo Universiteta*, ser. Fiziko-Matematicheskie Nauki 151(2), 36–44 (2009) (in Russian)
18. Voronenko, A. A., Chistikov, D. V.: On testing read-once Boolean functions in the basis B_5 . In: *Proceedings of the XVII International Workshop “Synthesis and complexity of control systems”*, pp. 24–30. Izdatel’sтво Instituta matematiki, Novosibirsk (2008) (in Russian)
19. Zubkov, O. V., Chistikov, D. V., Voronenko, A. A.: An upper bound on checking test complexity for almost all cographs. In: Wang, D. et al. (eds.) *13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2011)*, pp. 323–330. IEEE Computer Society, Los Alamitos (2012)